

0550
0408



OIPE

RAW SEQUENCE LISTING
PATENT APPLICATION: US/10/026,140
DATE: 04/08/2002
TIME: 09:42:41

Input Set : A:\GC697-SEQLIST.TXT
Output Set: N:\CRF3\04082002\J026140.raw

ENTERED

4 <110> APPLICANT: Dunn-Coleman, Nigel
5 Goedegebuur, Frits
6 Ward, Michael
7 Yao, Jian
9 <120> TITLE OF INVENTION: BGL5 Beta-Glucosidase and Nucleic Acids
10 Encoding the Same
12 <130> FILE REFERENCE: GC697
14 <140> CURRENT APPLICATION NUMBER: US 10/026,140
C--> 15 <141> CURRENT FILING DATE: 2002-03-26
17 <160> NUMBER OF SEQ ID NOS: 3
19 <170> SOFTWARE: FastSEQ for Windows Version 4.0
21 <210> SEQ ID NO: 1
22 <211> LENGTH: 1991
23 <212> TYPE: DNA
24 <213> ORGANISM: Trichoderma reesei
26 <400> SEQUENCE: 1
27 agccagggtcg caaccagcag cagcagcagc agtacagaga aatcaaccca gatagctcaa 60
28 aatgcccagag tcgctagctc tgcccaacga ctttgaatgg ggcttcgcaa cgcccgcccta 120
29 ccagatcgaa ggcgcgctca aagaagggtg cgcgcgcccg tccatctggg acacgtactg 180
30 ccacctggag ccacgcgcga ccaacggcgc caacggcgat gtggcttgcg atcactacca 240
31 ccgctacgat gaggactttg atctcttgac caagtaaggc gcaaaggcct accgcttctc 300
32 cttgtcgtgg tcgcgatca ttccctcgg cggcaggctg gatcccgta acgaggaggg 360
33 aattgagttt tacagcaaac tgattgacgc cctgttgagg cggggtatca cgccttgggg 420
34 gactttgtac cactgggata tgccacaggc gcttcacgat cgctatggag gctggctcaa 480
35 cgtggaagag gtccagctgg actttgagcg gtatgcgagg ttgtgctttg aacgttttgg 540
36 ggaccgagtc cagaactgga tcaccatcaa cgawccctgg attcaggcca tctatggata 600
37 tgccaccggc agcaacgccc cgggcaggag cagcattaac aagcactcca ccgagggcaa 660
38 cactgccact gagecgtggc tcgctggaaa ggcccagatc atgagccatg cccgcgcgct 720
39 ggccgtctac agcagggaact ttgcgccctc gcaaaagggc cagatcggca tctcgtcaa 780
40 cggcgactac tatgagccct gggacagcaa tgagcctcgg gacaaggagg ctgctgagcg 840
41 acgatggaa ttccacattg gctgggttgc caatcccatc ttcttgaaga aggactatcc 900
42 agagagcatg aagaagcagc tgggcgagag gcttcacgcc ctcaactccc cggactttgc 960
43 cctcctcaat gccggagaga ccgacttcta cggcatgaat tactacacat cccagttcgc 1020
44 gcgccaccta gacggtccc tcccggagac ggactatctc ggcgccatcc atgagacca 1080
45 ggagaataag gacggcagcc ccgttggcga ggagagcggc ctgcctggc tgcgctcctg 1140
46 cccggacatg ttccggaagc atctgcgccg ggtgtacggc ctgtacggca agcccatcta 1200
47 catcacggag aacggatgcc cgtgccctgg agaggagaac atgacgtgcg aggaggcgt 1260
48 caacgacccc ttccgcatcc gctactttga ctgcgacttg gactcgattt ccaaggccat 1320
49 taccagggac ggcgtcgtcg tcaaggggta ctttgcgtgg gcgttgctcg ataacttgg 1380
50 atggtcagat ggctacggac ccagattcgg cgtcacgttc acagactaca ccacctcaa 1440
51 gcgcacgccc aagaagtctg ccctggctct caaggacatg tttgcggccc ggcagagggt 1500
52 taaagtggcg gcataaagaa agggaaattt cttcttgcac tcagcctcta tgcattctcc 1560
53 tctctctttt tccctccctc cccttgcctc tctctctcta cctctcatat tccctctata 1620

RAW SEQUENCE LISTING

DATE: 04/08/2002

PATENT APPLICATION: US/10/026,140

TIME: 09:42:41

Input Set : A:\GC697-SEQLIST.TXT

Output Set: N:\CRF3\04082002\J026140.raw

```

54 ccccccgctt cttctcatga ccccatgctc cttgcccttg gcccctctct gtcgaattct 1680
55 gcctcttatac acgtcttatg cgtctgttta cttgcctttt ttttttttgt ctctttctgt 1740
56 ctgtctgtct gcctgtctat gtgtacctat ctggcccttc gtcattggc aacagatact 1800
57 agcacaagtt caagcaagca agcacgcaag caagcaagca agccagccat caacggcatc 1860
58 aaagcccat gtttagcctc atgttcacat tgctatgta tctacatcag ccattcacta 1920
59 ccaggcgaag aggccacaga gagtctcacc gtcttacctg tatatacgt tttttaaaaa 1980
60 aaaaaaaaaa a 1991
62 <210> SEQ ID NO: 2
63 <211> LENGTH: 484
64 <212> TYPE: PRT
65 <213> ORGANISM: Trichoderma reesei
67 <220> FEATURE:
68 <221> NAME/KEY: VARIANT
69 <222> LOCATION: (1)...(484)
70 <223> OTHER INFORMATION: Xaa = Any Amino Acid
72 <400> SEQUENCE: 2
73 Met Pro Glu Ser Leu Ala Leu Pro Asn Asp Phe Glu Trp Gly Phe Ala
74 1 5 10 15
75 Thr Ala Ala Tyr Gln Ile Glu Gly Ala Val Lys Glu Gly Gly Arg Gly
76 20 25 30
77 Pro Ser Ile Trp Asp Thr Tyr Cys His Leu Glu Pro Ser Arg Thr Asn
78 35 40 45
79 Gly Ala Asn Gly Asp Val Ala Cys Asp His Tyr His Arg Tyr Asp Glu
80 50 55 60
81 Asp Phe Asp Leu Leu Thr Lys Tyr Gly Ala Lys Ala Tyr Arg Phe Ser
82 65 70 75 80
83 Leu Ser Trp Ser Arg Ile Ile Pro Leu Gly Gly Arg Leu Asp Pro Val
84 85 90 95
85 Asn Glu Glu Gly Ile Glu Phe Tyr Ser Lys Leu Ile Asp Ala Leu Leu
86 100 105 110
87 Arg Arg Gly Ile Thr Pro Trp Val Thr Leu Tyr His Trp Asp Leu Pro
88 115 120 125
89 Gln Ala Leu His Asp Arg Tyr Gly Gly Trp Leu Asn Val Glu Glu Val
90 130 135 140
91 Gln Leu Asp Phe Glu Arg Tyr Ala Arg Leu Cys Phe Glu Arg Phe Gly
92 145 150 155 160
93 Asp Arg Val Gln Asn Trp Ile Thr Ile Asn Xaa Pro Trp Ile Gln Ala
94 165 170 175
95 Ile Tyr Gly Tyr Ala Thr Gly Ser Asn Ala Pro Gly Arg Ser Ser Ile
96 180 185 190
97 Asn Lys His Ser Thr Glu Gly Asn Thr Ala Thr Glu Pro Trp Leu Ala
98 195 200 205
99 Gly Lys Ala Gln Ile Met Ser His Ala Arg Ala Val Ala Val Tyr Ser
100 210 215 220
101 Arg Asp Phe Arg Pro Ser Gln Lys Gly Gln Ile Gly Ile Ser Leu Asn
102 225 230 235 240
103 Gly Asp Tyr Tyr Glu Pro Trp Asp Ser Asn Glu Pro Arg Asp Lys Glu
104 245 250 255
105 Ala Ala Glu Arg Arg Met Glu Phe His Ile Gly Trp Phe Ala Asn Pro

```

OK

RAW SEQUENCE LISTING

PATENT APPLICATION: US/10/026,140

DATE: 04/08/2002

TIME: 09:42:42

Input Set : A:\GC697-SEQLIST.TXT

Output Set: N:\CRF3\04082002\J026140.raw

```

106          260          265          270
107 Ile Phe Leu Lys Lys Asp Tyr Pro Glu Ser Met Lys Lys Gln Leu Gly
108          275          280          285
109 Glu Arg Leu Pro Ala Leu Thr Pro Ala Asp Phe Ala Ile Leu Asn Ala
110          290          295          300
111 Gly Glu Thr Asp Phe Tyr Gly Met Asn Tyr Tyr Thr Ser Gln Phe Ala
112 305          310          315          320
113 Arg His Leu Asp Gly Pro Val Pro Glu Thr Asp Tyr Leu Gly Ala Ile
114          325          330          335
115 His Glu His Gln Glu Asn Lys Asp Gly Ser Pro Val Gly Glu Glu Ser
116          340          345          350
117 Gly Leu Ala Trp Leu Arg Ser Cys Pro Asp Met Phe Arg Lys His Leu
118          355          360          365
119 Ala Arg Val Tyr Gly Leu Tyr Gly Lys Pro Ile Tyr Ile Thr Glu Asn
120          370          375          380
121 Gly Cys Pro Cys Pro Gly Glu Glu Asn Met Thr Cys Glu Glu Ala Val
122 385          390          395          400
123 Asn Asp Pro Phe Arg Ile Arg Tyr Phe Asp Ser His Leu Asp Ser Ile
124          405          410          415
125 Ser Lys Ala Ile Thr Gln Asp Gly Val Val Val Lys Gly Tyr Phe Ala
126          420          425          430
127 Trp Ala Leu Leu Asp Asn Leu Glu Trp Ser Asp Gly Tyr Gly Pro Arg
128          435          440          445
129 Phe Gly Val Thr Phe Thr Asp Tyr Thr Thr Leu Lys Arg Thr Pro Lys
130          450          455          460
131 Lys Ser Ala Leu Val Leu Lys Asp Met Phe Ala Ala Arg Gln Arg Val
132 465          470          475          480
133 Lys Val Ala Ala

```

136 <210> SEQ ID NO: 3

137 <211> LENGTH: 1455

138 <212> TYPE: DNA

139 <213> ORGANISM: Trichoderma reesei

141 <400> SEQUENCE: 3

```

142 atgcccgagt cgctagctct gcccaacgac tttgaatggg gcttcgcaac ggccgcctac      60
143 cagatcgaag gcgccgtcaa agaaggtggc cgcgcccggt ccattctgga cactactgc      120
144 cacctggagc catcgcgcac caacggcgcc aacggcgatg tggcttgcca tctactaccac      180
145 cgctacgatg aggactttga tctcttgacc aagtaaggcg caaaggccta ccgcttctcc      240
146 ttgtcgtggt cgcggatcat tcccctcggc ggcaggctgg atcccgtcaa cgaggaggga      300
147 attgagtttt acagcaaact gattgacgcc ctggttgagg ggggtatcac gccttggggtg      360
148 actttgtacc actgggatct gcctcaggcg cttcacgata gctatggagg ctggctcaac      420
149 gtggaagagg tccagctgga ctttgagcgg tatgagaggt tgtgctttga acgttttggg      480
150 gaccgagtcc agaactggat caccatcaac gawccctgga ttcaggccat ctatggatat      540
151 gccaccggca gcaacgcccc gggcaggagc agcattaaca agcactccac cgagggcaac      600
152 actgccactg agccgtggct cgctggaaag gccagatca tgagccatgc ccgcgcgctg      660
153 gccgtctaca gcagggaact tcgcccctcg caaaagggcc agatcggcct ctcgctcaac      720
154 ggcgactact atgagccctg ggacagcaat gagcctcggg acaaggaggc tgctgagcga      780
155 cggatggaat ttcacattgg ctggtttgcc aatcccatct tcttgaagaa ggactatcca      840
156 gagagcatga agaagcagct gggcgagagg cttccagccc tcaactccgc ggactttgcc      900
157 atcctcaatg ccggagagac cgacttctac ggcattgaatt actacacatc ccagttcgcg      960

```

RAW SEQUENCE LISTING

DATE: 04/08/2002

PATENT APPLICATION: US/10/026,140

TIME: 09:42:42

Input Set : A:\GC697-SEQLIST.TXT

Output Set: N:\CRF3\04082002\J026140.raw

```
158 cgccacctag acggtcccgt ccccgagacg gactatctcg gcgccatcca tgagcaccag 1020
159 gagaataagg acggcagccc cgttggcgag gagagcggcc tcgcctggct gcgctcctgc 1080
160 ccggacatgt tccggaagca tctcggccgg gtgtacggcc tgtacggcaa gcccatctac 1140
161 atcaccgaga acggatgccc gtgccctgga gaggagaaca tgacgtgcga ggaggccgctc 1200
162 aacgaccctt tccgcatccg stactttgac tcgcacttgg actcgatttc caaggccatt 1260
163 acccaggacg gcgtcgtcgt caaggggtac ttgctgtggg cgttgctcga taacttggaa 1320
164 tggtcagatg gctacggacc cagattcggc gtcacgttca cagactacac caccctcaag 1380
165 cgcacgcccc agaagtctgc cctggctcctc aaggacatgt ttgcggcccc gcagaggggtt 1440
166 aaagtggcgg cataa 1455
```

VERIFICATION SUMMARY

PATENT APPLICATION: US/10/026,140

DATE: 04/08/2002

TIME: 09:42:43

Input Set : A:\GC697-SEQLIST.TXT

Output Set: N:\CRF3\04082002\J026140.raw

L:15 M:271 C: Current Filing Date differs, Replaced Current Filing Date

L:93 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:2